A NOTE ON AVERAGES OVER GAUSSIAN RANDOM MATRIX ENSEMBLES

GABRIEL H. TUCCI

ABSTRACT. In this work we find a new formula for matrix averages over the Gaussian ensemble. Let $\mathbf{H}$ be an $n \times n$ Gaussian random matrix with complex, independent, and identically distributed entries of zero mean and unit variance. Given an $n \times n$ positive definite matrix $\mathbf{A}$, and a continuous function $f : \mathbb{R}^+ \to \mathbb{R}$ such that $\int_0^\infty e^{-\alpha t}|f(t)|^2\, dt < \infty$ for every $\alpha > 0$, we find a new formula for the expectation $\mathbb{E}[\mathrm{Tr}(f(\mathbf{HAH}^*))]$. Taking $f(x) = \log(1 + x)$ gives another formula for the capacity of the MIMO communication channel, and taking $f(x) = (1 + x)^{-1}$ gives the MMSE achieved by a linear receiver.

Random Matrices, Limiting Distribution, Gaussian Averages, MIMO Capacity, MMSE

1. INTRODUCTION

Random matrix theory was introduced to the theoretical physics community by Wigner in his work on nuclear physics in the 1950s ([23, 24]). Since that time, the subject is an important and active research area in mathematics and it finds applications in fields as diverse as the Riemann conjecture, physics, chaotic systems, multivariate statistics, wireless communications, signal processing, compressed sensing and information theory. In the last decades, a considerable amount of work has emerged in the communications and information theory on the fundamental limits of communication channels that makes use of results in random matrix theory [11, 19, 1]. For this reason, computing averages over certain matrix ensembles becomes extremely important in many situations. To be more specific, consider the well known case of the single user MIMO channel with multiple transmit and receive antennas. Denoting the number of transmitting antennas by t and the number of receiving antennas by r, the channel model is

$$\mathbf{y} = \mathbf{Hu} + \mathbf{n}$$

where $\mathbf{u} \in \mathbb{C}^t$ is the transmitted vector, $\mathbf{y} \in \mathbb{C}^r$ is the received vector, $\mathbf{H}$ is a $r \times t$ complex matrix and $\mathbf{n}$ is the zero mean complex Gaussian vector with independent, equal variance entries. We assume that $\mathbb{E}(\mathbf{nn}^*) = \mathbf{I}_r$,

where $(\cdot)^*$ denotes the complex conjugate transpose. It is reasonable to put a power constraint

$$\mathbb{E}(\mathbf{u}^*\mathbf{u}) = \mathbb{E}(\mathrm{Tr}(\mathbf{u}\mathbf{u}^*)) \leq P$$

where P is the total transmitted power. The signal to noise ratio, denoted by snr, is defined as the quotient of the signal power and the noise power and in this case is equal to P/r.

Recall that if $\mathbf{A}$ is an $n \times n$ Hermitian matrix then there exists $\mathbf{U}$ unitary and $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_n)$ such that $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^*$. Given a continuous function f we define $f(\mathbf{A})$ as

$$f(\mathbf{A}) = \mathbf{U}\mathrm{diag}(f(d_1), \ldots, f(d_n))\mathbf{U}^*.$$

Naturally, the simplest example is the one where $\mathbf{H}$ has independent and identically distributed (i.i.d.) Gaussian entries, which constitutes the canonical model for the single user narrow band MIMO channel. It is known that the capacity of this channel is achieved when $\mathbf{u}$ is a complex Gaussian zero mean and covariance $\mathrm{snr}\,\mathbf{I}_t$ vector (see for instance [20, 19]). For the fast fading channel, assuming statistical channel state information at the transmitter, the ergodic capacity is given by

$$\mathbb{E}\Big[\log\det(\mathbf{I}_r + \mathrm{snr}\mathbf{H}\mathbf{H}^*)\Big] = \mathbb{E}\Big[\mathrm{Tr}\log(\mathbf{I}_r + \mathrm{snr}\mathbf{H}\mathbf{H}^*)\Big] \qquad (1)$$

where in the last equality we use the fact that $\mathrm{Tr}\log(\cdot) = \log\det(\cdot)$. We refer the reader to [20] or [19] for more details on this.

Another important performance measure is the minimum mean square error (MMSE) achieved by a linear receiver, which determines the maximum achievable output signal to interference and noise ratio (SINR). For an input vector $\mathbf{x}$ with i.i.d. entries of zero mean and unit variance the MSE at the output of the MMSE receiver is given by

$$\min_{\mathbf{M}\in\mathbb{C}^{t\times r}} \mathbb{E}\Big[\|\mathbf{x} - \mathbf{M}\mathbf{y}\|^2\Big] = \mathbb{E}\Big[\mathrm{Tr}\Big(\mathbf{I}_t + \mathrm{snr}\mathbf{H}^*\mathbf{H}\Big)^{-1}\Big], \qquad (2)$$

where the expectation on the left hand side is over both the vectors $\mathbf{x}$ and the random matrices $\mathbf{H}$, while the right hand side is over $\mathbf{H}$ only. We refer to [19] for more details on this.

There is a big literature and history of work on averages over Gaussian ensembles; see for instance [20, 15, 19, 13, 6, 1, 11, 10, 16, 2, 9, 3] and references therein. In [20] the capacity of the Gaussian channel was computed as an improper integral. This integral is difficult to compute and asymptotic and simulation results are provided. In [3, 2, 13, 16, 17] several asymptotic results for large complex Gaussian random matrices are studied in connection with wireless communication and information theory. In [13] many aspects of correlated Gaussian matrices are addressed, in particular the capacity of Rayleigh channel was computed as the number of antennas increases to

infinity. The books [19, 11, 1] are excellent introductions to random matrix theory and their applications to physics and information theory. In [10] the spectral eigenvalue distribution for a random infinite d-regular graph was computed.

The typical approach in computing averages over random matrices is to consider the asymptotic behavior as the size of the matrix increases to infinity. In this work we contribute to this area by providing a unified framework to express the ergodic mutual information, the MSE at the output of the MMSE decoder and other types of functionals of a single user MIMO channel, when the number of transmitting and receiving antennas are equal and finite. We do not rely on asymptotic results as the number of antennas increases. The results shown in this work are new and novel to the best knowledge of the author and they were not discovered before.

In Section 2, we present some preliminaries in Schur polynomials that are later used in this work. In Section 3, we prove the main result of the paper, Theorem 3.2. This Theorem provides a new formula for the expectation

$$\mathbb{E}\Big[\mathrm{Tr}\Big(f(\mathbf{H}\mathbf{A}\mathbf{H}^*)\Big)\Big] \tag{3}$$

where $\mathbf{A}$ is positive definite matrix and f a continuous function such that

$$\int_0^\infty e^{-\alpha t}|f(t)|^2\,dt < \infty$$

for every $\alpha > 0$. Notice that, as previously stated, taking $f(x) = \log(1+x)$ gives another formula for the capacity of the MIMO communication channel, and taking $f(x) = (1+x)^{-1}$ gives the MMSE achieved by a linear receiver. We also discuss some applications and present some examples.

2. Schur Polynomials Preliminaries

A symmetric polynomial is a polynomial $P(x_1, x_2, \ldots, x_n)$ in n variables such that if any of the variables are interchanged one obtains the same polynomial. Formally, P is a symmetric polynomial if for any permutation σ of the set $\{1, 2, \ldots, n\}$ one has

$$P(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}) = P(x_1, x_2, \ldots, x_n).$$

Symmetric polynomials arise naturally in the study of the relation between the roots of a polynomial in one variable and its coefficients, since the coefficients can be given by a symmetric polynomial expressions in the roots. Symmetric polynomials also form an interesting structure by themselves. The resulting structures, and in particular the ring of symmetric functions, are of great importance in combinatorics and in representation theory (see for instance [4, 12, 8, 14] for more on details on this topic).
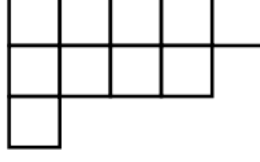
FIGURE 1. Young tableaux representation of the partition $(5, 4, 1)$.

The Schur polynomials are certain symmetric polynomials in n variables. This class of polynomials are very important in representation theory since they are the characters of irreducible representations of the general linear groups. The Schur polynomials are indexed by partitions. A partition of a positive integer n, also called an integer partition, is a way of writing n as a sum of positive integers. Two partitions that differ only in the order of their summands are considered to be the same partition. Therefore, we can always represent a partition λ of a positive integer n as a sequence of n non-increasing and non-negative integers d_i such that

$$\sum_{i=1}^{n} d_i = n \quad \text{with} \quad d_1 \geq d_2 \geq d_3 \geq \ldots \geq d_n \geq 0.$$

Notice that some of the d_i could be zero. Integer partitions are usually represented by the so called Young's tableaux (also known as Ferrers' diagrams). A Young tableaux is a finite collection of boxes, or cells, arranged in left-justified rows, with the row lengths weakly decreasing (each row has the same or shorter length than its predecessor). Listing the number of boxes on each row gives a partition λ of a non-negative integer n, the total number of boxes of the diagram. The Young diagram is said to be of shape λ, and it carries the same information as that partition. For instance, in Figure 1 we can see the Young tableaux corresponding to the partition $(5, 4, 1)$ of the number 10.

Given a partition λ of n

$$n = d_1 + d_2 + \cdots + d_n \quad : \quad d_1 \geq d_2 \geq \cdots \geq d_n \geq 0$$

the following functions are alternating polynomials (in other words they change sign under any transposition of the variables):

$$a_{(d_1,\ldots,d_n)}(x_1,\ldots,x_n) \;=\; \det \begin{bmatrix} x_1^{d_1} & x_2^{d_1} & \ldots & x_n^{d_1} \\ x_1^{d_2} & x_2^{d_2} & \ldots & x_n^{d_2} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{d_n} & x_2^{d_n} & \ldots & x_n^{d_n} \end{bmatrix}$$

$$= \sum_{\sigma \in S_n} \epsilon(\sigma) x_{\sigma(1)}^{d_1} \cdots x_{\sigma(n)}^{d_n}$$

FIGURE 2. Young tableaux representation of the partition $(5, 4, 1)$ with its corresponding hook's lengths.

where S_n is the permutation group of the set $\{1, 2, \ldots, n\}$. Since they are alternating, they are all divisible by the Vandermonde determinant

$$\Delta(x_1, \ldots, x_n) = \prod_{1 \leq j < k \leq n} (x_j - x_k).$$

The Schur polynomial associated to λ is defined as the ratio:

$$s_\lambda(x_1, x_2, \ldots, x_n) = \frac{a_{(d_1+n-1, d_2+n-2, \ldots, d_n+0)}(x_1, \ldots, x_n)}{\Delta(x_1, \ldots, x_n)}.$$

This is a symmetric function because the numerator and denominator are both alternating, and a polynomial since all alternating polynomials are divisible by the Vandermonde determinant (see [4, 8, 14] for more details here). For instance,

$$s_{(2,1,1)}(x_1, x_2, x_3) = x_1 \, x_2 \, x_3 \, (x_1 + x_2 + x_3)$$

and

$$\begin{aligned} s_{(2,2,0)}(x_1, x_2, x_3) &= x_1^2 \, x_2^2 + x_1^2 \, x_3^2 + x_2^2 \, x_3^2 + x_1^2 \, x_2 \, x_3 \\ &+ x_1 \, x_2^2 \, x_3 + x_1 \, x_2 \, x_3^2. \end{aligned}$$

Another definition we need for the next Section is the so called hook length, $\text{hook}(x)$, of a box x in Young diagram of shape λ. This is defined as the number of boxes that are in the same row to the right of it plus those boxes in the same column below it, plus one (for the box itself). For instance, in Figure 2 we can see the hook lengths of the partition $(5, 4, 1)$. The product of the hook's length of a partition is the product of the hook lengths of all the boxes in the partition.

We recommend the interested reader to consult [4, 8, 14] for more details and examples on this topic.

3. Averages over Gaussian Ensembles

Let M_n be the set of all $n \times n$ complex matrices and $\mathbf{U}_n$ the set of $n \times n$ unitary complex matrices. Let $d\mathbf{H}$ be the Lebesgue measure on M_n and let

$$d\nu(\mathbf{H}) = \pi^{-n^2} \exp\Big(-\operatorname{trace}(\mathbf{H}^*\mathbf{H})\Big) d\mathbf{H}$$

be the Gaussian measure on M_n. This is the induced measure by the Gaussian random matrix with complex independent and identically distributed entries with zero mean and unit variance in the set of matrices, when this is represented as an Euclidean space of dimension $2n^2$. Note that this probability measure is left and right invariant under unitary multiplication (i.e., $d\nu(\mathbf{H}\mathbf{U}) = d\nu(\mathbf{U}\mathbf{H}) = d\nu(\mathbf{H})$ for every unitary $\mathbf{U}$). The following Theorem can be found in page 447 of [8].

Theorem 3.1. [8] *For every* $\mathbf{A}, \mathbf{B}$ *Hermitian* $n \times n$ *matrices and every partition* λ

$$\int_{M_n} s_\lambda(\mathbf{A}\mathbf{H}^*\mathbf{B}\mathbf{H}) \, d\nu(\mathbf{H}) = h(\lambda) s_\lambda(\mathbf{A}) s_\lambda(\mathbf{B}) \tag{4}$$

where $h(\lambda)$ *is the product of the hook–lengths of* λ.

Denote by $(m-k, 1^k)$ the partition $(m-k, 1, 1, \ldots, 1)$ with k ones. It is a well known fact in matrix theory (see [4] or [8]) that for every Hermitian $n \times n$ matrix $\mathbf{A}$ and for every integer m

$$\operatorname{Tr}(\mathbf{A}^m) = \sum_{k=0}^{n-1} (-1)^k s_{(m-k,1^k)}(\mathbf{A}). \tag{5}$$

Note that for the case $1 \leq m < n$, even though the sum is up to the $n-1$ term, all the terms between $\min\{n, m\}$ and $n-1$ are zero. In particular,

- $\operatorname{Tr}(\mathbf{A}) = s_{(1)}(\mathbf{A})$
- $\operatorname{Tr}(\mathbf{A}^2) = s_{(2,0)}(\mathbf{A}) - s_{(1,1)}(\mathbf{A})$
- $\operatorname{Tr}(\mathbf{A}^3) = s_{(3,0)}(\mathbf{A}) - s_{(2,1)}(\mathbf{A}) + s_{(1,1,1)}(\mathbf{A})$
- $\operatorname{Tr}(\mathbf{A}^4) = s_{(4,0)}(\mathbf{A}) - s_{(3,1)}(\mathbf{A}) + s_{(2,1,1)}(\mathbf{A}) - s_{(1,1,1,1)}(\mathbf{A})$.

The constant $s_{(m-k,1^k)}(\mathbf{I}_p)$ is equal to

$$s_{(m-k,1^k)}(\mathbf{I}_p) = \frac{(m+p-(k+1))!}{k!(p-(k+1))!(m-(k+1)!m}$$

(see [8] for a proof of this formula). Therefore,

$$\frac{s_{(m-k,1^k)}(\mathbf{I}_p)}{s_{(m-k,1^k)}(\mathbf{I}_n)} = \frac{(m+p-(k+1))!}{(m+n-(k+1))!} \cdot \frac{(n-(k+1))!}{(p-(k+1))!}. \tag{6}$$

For every $\alpha > 0$ let us define the following class of functions

$$L^2_\alpha : \quad = \quad \Big\{ f : \mathbb{R}^+ \to \mathbb{R} : \text{measurable and such that}$$

$$\int_0^\infty e^{-\alpha t} \, |f(t)|^2 \, dt < \infty \Big\}.$$

This is a Hilbert space with respect to the inner product

$$\langle f, g \rangle_\alpha = \int_0^\infty e^{-\alpha t} \, f(t) g(t) \, dt.$$

Moreover, polynomials are dense with respect to this norm (see Chapter 10 in [18]). Let $\mathcal{A}_\alpha$ be the set of continuous functions in L^2_α and let $\mathcal{A}$ be the intersection of all the $\mathcal{A}_\alpha$,

$$\mathcal{A} = \cap_{\alpha > 0} \mathcal{A}_\alpha.$$

Note that the family $\mathcal{A}$ is a very rich family of functions. For instance, all functions that do not grow faster than polynomials belong to these family. In particular, $f(t) = \log(1 + t) \in \mathcal{A}$.

Theorem 3.2. *Let* $\mathbf{A}$ *be an* $n \times n$ *positive definite matrix and let* $\{d_1, \ldots, d_n\}$ *be the set of eigenvalues of* $\mathbf{A}$*. Assume that all the eigenvalues are different. Then for every* $f \in \mathcal{A}$ *we have that*

$$\int_{M_n} \mathrm{Tr}\Big(f(\mathbf{H}^* \mathbf{A} \mathbf{H})) \Big) \, d\nu(\mathbf{H}) = \frac{1}{\det(\Delta(\mathbf{D}))} \sum_{k=0}^{n-1} \det(\mathbf{T}_k) \tag{7}$$

where $\Delta(\mathbf{D})$ *is the Vandermonde matrix associated with the matrix* $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_n)$ *and* $\mathbf{T}_k$ *is the matrix constructed by replacing the* $(k+1)$ *row of* $\Delta(\mathbf{D})$ *(*$\{d_i^{n-(k+1)}\}_{i=1}^n$*) by*

$$\frac{1}{(n - (k+1))!} \{f_k(d_i)\}_{i=1}^n$$

where

$$f_k(x) := \int_0^\infty e^{-t} (tx)^{n-(k+1)} f(tx) \, dt.$$

Proof. First, we will prove the Theorem for polynomials. Let p and q be two polynomials. It is clear that

$$\mathrm{Tr}\big((p + q)(\mathbf{H}^* \mathbf{A} \mathbf{H}) \big) = \mathrm{Tr}\big(p(\mathbf{H}^* \mathbf{A} \mathbf{H}) \big) + \mathrm{Tr}\big(q(\mathbf{H}^* \mathbf{A} \mathbf{H}) \big)$$

and $(p + q)_k = p_k + q_k$ for every $k = 0, \ldots, n-1$. Therefore, both sides of the Equation (7) are linear and it is enough to prove the Theorem for the case $p(x) = x^m$ with $m \geq 0$. Using Theorem 3.1 and Equation (5) we see that for every positive definite $n \times n$ matrix $\mathbf{A}$, the average

$$\int_{M_n} \mathrm{Tr}\Big((\mathbf{H}^* \mathbf{A} \mathbf{H})^m \Big) \, d\nu(\mathbf{H})$$

is equal to

$$\sum_{k=0}^{n-1} (-1)^k h(\lambda_k) s_{\lambda_k}(\mathbf{A}) s_{\lambda_k}(\mathbf{I}_n)$$

where λ_k is the partition $(m-k, 1^k)$.

It is well known (see [4]) that for every partition $\lambda = (\lambda_1, \ldots, \lambda_n)$

$$s_\lambda(\mathbf{I}_n) = \prod_{1 \le i \le j \le n} \frac{\lambda_i - \lambda_j + j - i}{j - i}. \tag{8}$$

Therefore, we can deduce that

$$s_{\lambda_k}(\mathbf{I}_n) = \frac{1}{m} \cdot \frac{(m+n-(k+1))!}{k!\,(n-(k+1))!\,(m-(k+1))!}. \tag{9}$$

We can see by direct examination that the hook–length of the partition λ_k is equal to

$$h(\lambda_k) = k!\,(m-(k+1))!\,m.$$

Hence,

$$s_{\lambda_k}(\mathbf{I}_n) h(\lambda_k) = \frac{(m+n-(k+1))!}{(n-(k+1))!}.$$

Since $\mathbf{A}$ is a positive definite matrix, by the spectral Theorem there exists $\mathbf{U}$ unitary and $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_n)$ diagonal such that $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^*$. Note that the d_i are the eigenvalues of $\mathbf{A}$. By definition of the Schur polynomials

$$s_{\lambda_k}(\mathbf{A}) = s_{\lambda_k}(\mathbf{D}) = \frac{\det(\mathbf{S}_k)}{\det(\Delta(\mathbf{D}))}$$

where $\Delta(\mathbf{D})$ is the Vandermonde matrix associated with the sequence $\{d_i\}_{i=1}^n$ and $\mathbf{S}_k$ is a matrix whose i–th column is equal to

$$\begin{pmatrix} d_i^{n-1+m-k} \\ d_i^{n-2+1} \\ d_i^{n-3+1} \\ \vdots \\ d_i^{n-(k+1)+1} \\ d_i^{n-(k+2)} \\ \vdots \\ d_i^{n-(n-1)} \\ 1 \end{pmatrix}.$$

It is easy to see that after k transpositions of the rows of the matrix $\mathbf{S}_k$ we obtain a new matrix $\mathbf{H}_k$ whose i–th column is equal to

$$
\begin{pmatrix}
d_i^{n-1} \\
d_i^{n-2} \\
d_i^{n-3} \\
\vdots \\
d_i^{n-k} \\
d_i^{n+m-(k+1)} \\
d_i^{n-(k+2)} \\
\vdots \\
d_i^{n-(n-1)} \\
1
\end{pmatrix}.
$$

This matrix is equal to the Vandermonde matrix $\Delta(\mathbf{D})$ except for the $(k+1)$ row, $\{d_i^{n-(k+1)}\}_{i=1}^n$, which is substituted by the row $\{d_i^{n+m-(k+1)}\}_{i=1}^n$. Note also that

$$
\det(\mathbf{S}_k) = (-1)^k \det(\mathbf{H}_k).
$$

Therefore, the average

$$
\int_{M_n} \mathrm{Tr}\Big((\mathbf{H}^*\mathbf{A}\mathbf{H})^m\Big)\, d\nu(\mathbf{H})
$$

is equal to

$$
\frac{1}{\det(\Delta(\mathbf{D}))} \sum_{k=0}^{n-1} \frac{(m+n-(k+1))!}{(n-(k+1))!} \cdot \det(\mathbf{H}_k).
$$

Using the fact that $\int_0^\infty e^{-t} t^p\, dt = p!$ and the definition of $p_k(x)$ for the case $p(x) = x^m$ we see that

$$
\begin{aligned}
p_k(x): &= \int_0^\infty e^{-t}(tx)^{n+m-(k+1)}\, dt \\
&= (m+n-(k+1))!\; x^{m+n-(k+1)}.
\end{aligned}
$$

Therefore, our claim holds and we have proven the result for all polynomials. Now consider $f \in \mathcal{A}$ and let β be the maximum eigenvalue, i.e., $\beta = \max\{d_1, \ldots, d_n\}$. Define $\alpha = 1/\beta$. Since $f \in \mathcal{A}$, then $f \in \mathcal{A}_\alpha$ and let $\{p^{(r)}\}_{r\geq 1}$ be a sequence of polynomials such that $\|f - p^{(r)}\|_\alpha \to 0$. Let $\mathbf{T}_k^{(n)}$ be the matrix constructed by replacing the $(k+1)$ row of $\Delta(\mathbf{D})$ ($\{d_i^{n-(k+1)}\}_{i=1}^n$) by

$$
\frac{1}{(n-(k+1))!}\{p_k^{(r)}(d_i)\}_{i=1}^n
$$

where

$$
p_k^{(r)}(x) := \int_0^\infty e^{-t}(tx)^{n-(k+1)} p^{(r)}(tx)\, dt.
$$

Let $\mathbf{T}_k$ be the matrix constructed by replacing the $(k+1)$ row of $\Delta(\mathbf{D})$ by

$$\frac{1}{(n-(k+1))!}\{f_k(d_i)\}_{i=1}^n$$

where

$$f_k(x) := \int_0^\infty e^{-t}(tx)^{n-(k+1)}f(tx)\,dt.$$

To prove that Equation (7) holds it is enough to prove that

$$\det(\mathbf{T}_k^{(n)}) \to \det(\mathbf{T}_k)$$

as $n \to \infty$ for every $k = 0, 1, \ldots, n-1$. For this, it is enough to prove that $p_k^{(r)}(d_i) \to f_k(d_i)$ for every k and every $i = 1, 2, \ldots, n$. Note that

$$
\begin{aligned}
|f_k(d_i) &- p_k^{(r)}(d_i)| \\
&= \int_0^\infty e^{-t}(td_i)^{n-(k+1)}|f(td_i) - p^{(r)}(td_i)|\,dt \\
&\leq d_i^{n-(k+1)}\sqrt{(2(n-(k+1)))!} \cdot \\
&\quad \left(\int_0^\infty e^{-t}|f(td_i) - p^{(r)}(tdi)|^2\,dt\right)^{\frac{1}{2}} \\
&= d_i^{n-(k+\frac{3}{2})}\sqrt{(2(n-(k+1)))!} \cdot \\
&\quad \left(\int_0^\infty e^{-\frac{t}{d_i}}|f(t) - p^{(r)}(t)|^2\,dt\right)^{\frac{1}{2}}
\end{aligned}
$$

where we use Cauchy-Schwartz for the second inequality and change of variable for the last one. Now, by construction the sequence $\{p^{(r)}\}$ satisfies

$$\lim_{n\to\infty}\|f - p^{(r)}\|_\alpha^2 = \lim_{n\to\infty}\int_0^\infty e^{-\alpha t}|f(t) - p^{(r)}(t)|^2\,dt = 0$$

and $\alpha \leq d_i^{-1}$. Hence, we see that

$$\lim_{n\to\infty}|f_k(d_i) - p_k^{(r)}(d_i)| = 0$$

finishing the proof. $\square$

Remark 3.3. *We would like to observe that the case when not all the eigenvalues are different can be treated as above by perturbing of the original eigenvalues and applying a subsequent limit. We present an instance of this situation in Corollary 3.6.*

As a consequence we have a new formula for the capacity of the MIMO communication channel and for the MMSE described in the introduction.

Corollary 3.4. *Let* $\mathbf{A}$ *be as in Theorem 3.2. Then*

$$\int_{M_n} \mathrm{Tr}\Big(\log(\mathbf{I}_n + \mathbf{H}^*\mathbf{A}\mathbf{H}) \Big) \, d\nu(\mathbf{H})$$

is equal to

$$\frac{1}{\det(\Delta(\mathbf{D}))} \sum_{k=0}^{n-1} \det(\mathbf{T}_k) \tag{10}$$

where $\mathbf{T}_k$ *is the matrix constructed by replacing the* $(k+1)$ *row of* $\Delta(\mathbf{D})$
$(\{d_i^{n-(k+1)}\}_{i=1}^n)$ *by*

$$\left\{ \frac{1}{(n-(k+1))!} \int_0^\infty e^{-t}(td_i)^{n-(k+1)} \log(1+td_i) \, dt \right\}_{i=1}^n .$$

Corollary 3.5. *Let* $\mathbf{A}$ *be as in theorem 3.2. Then*

$$\int_{M_n} \mathrm{Tr}\Big((\mathbf{I}_n + \mathbf{H}^*\mathbf{A}\mathbf{H})^{-1} \Big) \, d\nu(\mathbf{H})$$

is equal to

$$\frac{1}{\det(\Delta(\mathbf{D}))} \sum_{k=0}^{n-1} \det(\mathbf{T}_k) \tag{11}$$

where $\mathbf{T}_k$ *is the matrix constructed by replacing the* $(k+1)$ *row of* $\Delta(\mathbf{D})$
$(\{d_i^{n-(k+1)}\}_{i=1}^n)$ *by*

$$\left\{ \frac{1}{(n-(k+1))!} \int_0^\infty e^{-t}(td_i)^{n-(k+1)}(1+td_i)^{-1} \, dt \right\}_{i=1}^n .$$

As an application let us compute explicitly the two dimensional case for the capacity.

Corollary 3.6. *Let* $\mathbf{A}$ *be an Hermitian* 2×2 *matrix with eigenvalues* d_1
and d_2. *If* $d_1 \neq d_2$ *then*

$$\int_{M_2} \mathrm{Tr}\Big(\log(\mathbf{I}_2 + \mathbf{H}^*\mathbf{A}\mathbf{H}) \Big) \, d\nu(\mathbf{H})$$

is equal to

$$\frac{f_0(d_1) - f_0(d_2) + d_1 f_1(d_2) - d_2 f_1(d_1)}{d_1 - d_2}$$

where $f_0(d_i) = \int_0^\infty e^{-t} t d_i \log(1+td_i) \, dt$ *and* $f_1(d_i) = \int_0^\infty e^{-t} \log(1+td_i) \, dt$.
If $d_1 = d_2 = d$ *then*

$$\int_{M_2} \mathrm{Tr}\Big(\log(\mathbf{I}_2 + d \cdot \mathbf{H}^*\mathbf{H}) \Big) \, d\nu(\mathbf{H})$$

is equal to

$$\int_0^\infty e^{-t}\left[(1+t)\log(1+td) + \frac{td(t-1)}{1+td}\right]dt.$$

Proof. The case $d_1 \neq d_2$ is a direct application of theorem 3.2 for $n = 2$ and $f(x) = \log(1+x)$. For the case $d_1 = d_2 = d$ then both the top and the bottom vanish and we have to take the limit of $d_1 = d + \epsilon$ and $d_2 = d$ as $\epsilon \to 0$. More precisely,

$$\lim_{\epsilon \to 0} \frac{f_0(d+\epsilon) - f_0(d)}{\epsilon} = \int_0^\infty e^{-t}\left[t\log(1+td) + \frac{t^2 d}{1+td}\right]dt$$

and

$$\lim_{\epsilon \to 0} \frac{(d+\epsilon)f_1(d) - df_1(d+\epsilon)}{\epsilon} =$$
$$\int_0^\infty e^{-t}\left[(1+d)\log(1+td) - \frac{td}{1+td}\right]dt.$$

Putting all the pieces together we finish the proof. $\square$

Analogously, we can compute explicitly the moments for the two dimensional case.

Theorem 3.7. *Let* $\mathbf{A}$ *be an Hermitian* 2×2 *matrix with eigenvalues* d_1 *and* d_2 *and let* $m \geq 1$. *If* $d_1 \neq d_2$ *then*

$$\int_{M_2} \mathrm{Tr}\left((\mathbf{H}^*\mathbf{A}\mathbf{H})^m\right) d\nu(\mathbf{H})$$

is equal to

$$m!\left((m+1)\frac{d_1^{m+1} - d_2^{m+1}}{d_1 - d_2} + \frac{d_1 d_2^m - d_2 d_1^m}{d_1 - d_2}\right).$$

If $d_1 = d_2 = d$ *then*

$$\int_{M_2} \mathrm{Tr}\left((\mathbf{H}^*\mathbf{A}\mathbf{H})^m\right) d\nu(\mathbf{H}) = m!\,(m^2 + m + 2)d^m.$$

4. Conclusion

Using results on random matrix theory and representation theory, in particular Schur polynomials, we prove a new formula for the average of functionals over the Gaussian ensemble. In particular, this gives another formula for the capacity of the MIMO Gaussian channel and the MMSE achieved by a linear receiver.

References

[1] Anderson G., Guionnet A. and Zeitouni O., *An Introduction to Random Matrices*, Cambridge University Press, 2009.

[2] Debbah M., Hachem W., Loubaton P. and de Courville M., *MMSE analysis of certain large isometric random precoded systems*, IEEE Trans. on Information Theory, vol. 49, pp. 1293-1311, 2003.

[3] Edelman A., *Eigenvalues and condition numbers of random matrices*, Ph.D. thesis, Dept. Mathematics, MIT, 1989.

[4] Fulton H., *Representation Theory*, Springer, 1991.

[5] Fyodorov Y.V. and Khoruzhenko B.A., *A few remarks on Colour–Flavour Transformations, truncation of random unitary matrices, Berezin reproducing kernels and Selberg type integrals*, arXiv : math-ph/0610045v2, 2006.

[6] Marzetta T.L. and Hochwald B.M., *Capacity of a Mobile Multiple-Antenna Communication Link in Rayleigh Flat Fading*, IEEE Trans. Inform. Theory, vol. 45, p. 139-157, Jan. 1999.

[7] Marzetta T., Tucci G. and Simon S., *A Random Matrix–Theoretic Approach to Handling Singular Covariance Estimates*, submitted to IEEE Trans. Inf. Theory, 2010.

[8] Macdonald I.G., *Symmetric functions and Hall Polynomials*, Clarendon Press, Oxford University Press, New York, 1995.

[9] Marchenko V. and Pastur L., *Distribution of eigenvalues for some sets of random matrices*, Math. USSR-Sbornik, vol. 1, pp. 457-483, 1967.

[10] McKay B., *The Expected Eigenvalue Distribution of a Large Regular Graph*, Lin. Alg. and Applications, vol. 40, pp. 203-216, 1981.

[11] Mehta M., *Random Matrices*, Academic Press, vol. 142, Third Edition, 2004.

[12] Muirhead R.J., *Aspects of Multivariate Statistical Theory*, John Wiley & Sons, New York, 1982.

[13] Ratnarajah, Vaillancourt, Alvo, *Complex random matrices and Rayleigh channel capacity*, Commun. Inf. Syst., vol. 3, no. 2, pp. 119-138, 2003.

[14] Sagan B., *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions*, Springer, 2nd edition, 2010.

[15] Simon S.H., Moustakas A.L., Marinelli, L., *Capacity and Character Expansions: Moment-Generating Function and Other Exact Results for MIMO Correlated Channels*, IEEE Transactions on Information Theory, vol. 52, no. 12, pp. 5336-5351, 2006.

[16] Silverstein J. and Bai Z., *On the empirical distribution of eigenvalues of a class of large dimensional random matrices*, J. Multivariate Anal., vol. 54 , pp. 175-192, 1995.

[17] Silverstein J. and Choi S., *Analysis of the limiting spectral distribution of large dimensional random matrices*, J. Multivariate Anal., vol. 54, pp. 295-309, 1995.

[18] Spain B., Smith M., *Functions of Mathematical Physics*, von Nostrand Reinhold company, 1970.

[19] Verdu. S and Tulino A.M., *Random Matrix Theory and Wireless Communications*, Now Publishers Inc., 2004.

[20] Telatar E., *Capacity of Multi–antenna Gaussian Channels*, European Trans. on Telecommunications, vol. 10, no. 6, pp. 585-596, 1999.

[21] D. Voiculescu, *Free Probability Theory*, Fields Institute Communications, 1997 .

[22] D. Voiculescu, K. Dykema and A. Nica, *Free Random Variables*, CRM Monograph Series, vol. 1, AMS, 1992.

[23] Wigner E., *Characteristic vectors of bordered matrices with infinite dimensions*, Annals of Mathematics, vol. 62, pp. 546-564, 1955.

[24] Wigner E., *On the distribution of roots of certain symmetric matrices*, Annals of Mathematics, vol. 67, pp. 325-327, 1958.